

Dissociative conceptual and quantitative problem solving outcomes across interactive engagement and traditional format introductory physics

Mark A. McDaniel,^{1,2,*} Siera M. Stoen,³ Regina F. Frey,^{2,4,5} Zachary E. Markow,⁶
K. Mairin Hynes,³ Jiuqing Zhao,² and Michael J. Cahill²

¹*Department of Psychological and Brain Sciences, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

²*Center for Integrative Research on Cognition, Learning, and Education, Washington University in St. Louis,
One Brookings Drive, Saint Louis, Missouri 63130, USA*

³*Department of Physics, Washington University in St. Louis, One Brookings Drive, Saint Louis,
Missouri 63130, USA*

⁴*Department of Chemistry, Washington University in St. Louis, One Brookings Drive, Saint Louis,
Missouri 63130, USA*

⁵*The Teaching Center, Washington University in St. Louis, One Brookings Drive, Saint Louis,
Missouri 63130, USA*

⁶*Department of Biomedical Engineering, Washington University in St. Louis, One Brookings Drive,
Saint Louis, Missouri 63130, USA*

(Received 1 June 2016; published 22 November 2016)

The existing literature indicates that interactive-engagement (IE) based general physics classes improve conceptual learning relative to more traditional lecture-oriented classrooms. Very little research, however, has examined quantitative problem-solving outcomes from IE based relative to traditional lecture-based physics classes. The present study included both pre- and post-course conceptual-learning assessments and a new quantitative physics problem-solving assessment that included three representative conservation of energy problems from a first-semester calculus-based college physics course. Scores for problem translation, plan coherence, solution execution, and evaluation of solution plausibility were extracted for each problem. Over 450 students in three IE-based sections and two traditional lecture sections taught at the same university during the same semester participated. As expected, the IE-based course produced more robust gains on a Force Concept Inventory than did the lecture course. By contrast, when the full sample was considered, gains in quantitative problem solving were significantly greater for lecture than IE-based physics; when students were matched on pre-test scores, there was still no advantage for IE-based physics on gains in quantitative problem solving. Further, the association between performance on the concept inventory and quantitative problem solving was minimal. These results highlight that improved conceptual understanding does not necessarily support improved quantitative physics problem solving, and that the instructional method appears to have less bearing on gains in quantitative problem solving than does the kinds of problems emphasized in the courses and homework and the overlap of these problems to those on the assessment.

DOI: [10.1103/PhysRevPhysEducRes.12.020141](https://doi.org/10.1103/PhysRevPhysEducRes.12.020141)

I. INTRODUCTION

In the past 20 years there has been extensive interest in the physics education literature regarding the use of interactive-engagement (IE) techniques (also termed “active learning”) to improve student learning in introductory physics courses relative to traditional lecture curricula

[1,2,3], as well as in introductory STEM courses in general [4,5]. IE activities are focused on promoting conceptual understanding and problem solving primarily through active participation, and are often stimulated by techniques such as cooperative group problem solving [6], predicting outcomes for classroom demonstrations [7], short conceptual problems or “understanding checks” [2] followed by discussion and feedback, and contextually situated (i.e., real-world) problem scenarios for which students must assemble (through discussion and group work) appropriate physics principles toward a solution.

A growing evidence base has reinforced the conclusion that conceptual-learning outcomes are improved through IE-enriched classrooms relative to more traditional lecture-oriented classrooms. With regard to introductory physics,

*Corresponding author.

Department of Psychological and Brain Sciences, Washington University, One Brookings Drive, Saint Louis, Missouri 63130, USA.

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

quasiexperimental studies within a university-specific course [1,2] and survey studies comparing across institutions [3] have almost without exception reported that IE-rich instruction produces superior student outcomes in terms of conceptual understanding than do traditional lecture courses. A key issue in physics instruction that has received less investigation, however, concerns the assumption that improved conceptual understanding (as achieved through IE courses) should support improved physics problem solving [8]. The objective of the present study was to explore whether an IE course, a course that we had previously shown to produce superior gains in conceptual understanding (relative to a traditional course) [1], supported corresponding gains in physics problem-solving performance (relative to a traditional course), using quantitative physics problems that were sampled from a university-level first year physics course.

Before continuing, we emphasize that the quasiexperiment reported here was not designed to isolate a particular component of the instructional context as causing specific outcomes. We assume that instructional outcomes are influenced by the content emphasized during class (*what* is taught, i.e., the curriculum) [9], by the instructional methods used to assist the students in learning the content (*how* the content is taught, e.g., by active learning techniques), and by the practice that is required of students (e.g., homework assignments) [10]. All of these components differed across the IE course and traditional courses that we examined (similar to the preponderance of extant research in this arena). The IE course not only included interactive-engagement class techniques (by contrast, the traditional course relied primarily on lecture), but the IE course covered somewhat different content than the traditional course. Generally, active learning in IE courses focuses on confronting students with “conceptual” problems that do not require quantitative problem solving, and quantitative problems that are situated in context-rich scenarios (at least in the IE physics course investigated herein). By contrast, conceptual problems are largely absent from traditional physics courses (at least in the current traditional course), and much of class time is spent on deriving formulas and solutions for multistep quantitative problems. We addressed the issue of whether the conceptual focus of an IE-rich introductory physics class provides benefits to learning to solve quantitative physics problems relative to a more traditional (lecture-based) physics class.

A. Interactive engagement courses and quantitative problem solving

Teaching students to solve quantitative physics problems is a central objective in physics instruction (at least for math-based courses), yet the extant research on the outcomes of IE-rich instruction includes very few investigations of students’ quantitative problem-solving performance. For instance, a recent meta-analysis that was

conducted across a range of introductory STEM courses identified 21 published studies contrasting physics and astronomy courses taught using IE-rich instruction versus traditional lecture instruction [4]. To evaluate outcomes, most of these studies used concept inventories requiring little or no quantitative problem solving (e.g., on the Force Concept Inventory, the only numbers are given as fractions of an original value; that is, the only computation would be to figure out what is the new force if a particle is moved twice as far away). Accordingly, these results do not generally inform the question of whether IE-rich courses produce superior outcomes on students’ ability to solve quantitative physics problems. Only four of the published studies explicitly indicated using quantitative problem-solving outcomes; we review those next.

Hake [3] contrasted outcomes from IE engagement courses with traditional courses (relying heavily on lecture), examining a range of high school to university introductory physics courses. This study did include the mechanics baseline (MB) test, a test that requires problem solving based on “conceptual understanding in addition to some mathematical skill” (p. 68). The Hake survey indicated that IE courses generally were associated with higher MB scores than were traditional courses. Similarly, Crouch, and Mazur [11] reported modest (4%) improvement in MB scores for an active learning course (“peer instruction” approach) relative to a traditional course taught the preceding year (no statistical comparison was provided). In addition, significantly better performance was observed for the active learning course on an exam of quantitative problems relative to a traditional course taught six years previously.

Thacker *et al.* [12] examined performance on a single two-part quantitative physics problem (electrical circuit problem) added to the course exams in an inquiry-based physics course (with 24 students) and 3 traditionally taught courses (one enrolled physics honors students, one enrolled primarily engineering students, and one enrolled non-science majors). Credit for the problem solution to each part was given only for a fully correct solution (no partial credit given). The results were mixed, with the inquiry students performing marginally or significantly better on the problem than the “engineering” and non-science sections, respectively, but significantly more poorly than the honors section. Hoellwarth *et al.* [13] contrasted active-learning instruction implemented in “studio” courses with comparably sized (about 40 students) traditional-lecture instruction in introductory calculus-based college physics courses over the course of three years. Performance on 4–5 quantitative final exam problems was assessed. The results were again mixed: In two of the years traditional instruction produced better performance than active-learning instruction (significantly so in one year but not the other), whereas in a third year performance was somewhat higher

(nonsignificantly) for the active-learning based instruction than lecture instruction.

Thus, the few existing studies show mixed results, and importantly several methodological issues limit the conclusiveness of these studies. First, quantitative problem solving assessments in these studies were given primarily as a post-test (after varying weeks of classroom instruction); that is, a precourse physics problem-solving assessment was not administered. Further, even when indices of students' incoming ability were available (none was available for several studies), these indices were not included as covariates in statistical analyses. Thus, one cannot determine if differences in students' physics problem solving were present at the outset across the different course formats (IE vs traditional; this was the case even in Hoellwarth *et al.* [13] who established that students in the different course formats had comparable math SAT, verbal SAT, and precourse conceptual inventory scores). Such differences could exist, especially in the Thacker *et al.* [12] study in which the different classes were explicitly enrolled with students of different backgrounds, and perhaps in Crouch and Mazur [11], where the advantage of IE relative to traditional physics on an exam of quantitative problems emerged when the two courses were taught six years apart. Without taking possible precourse differences of students (if present) into account, the extant results are not decisive [5].

Second, there is concern that the problems used in these studies were somewhat elementary. For instance, some have suggested that problems on the MB do not adequately assess more advanced abilities required for problems that are standard in physics courses [3]. Similarly, the electrical circuit problem in the Thacker *et al.* [12] study was relatively simple, requiring two simple equations (one for adding resistors in series, and one for adding resistors in parallel), and some instructors *a priori* might classify this as a conceptual problem. Indeed, performance on the MB is very highly correlated with performance on the FCI ($r > 0.91$) [3], so arguably the two instruments are in fact capturing a single latent construct, that of conceptual understanding. Of course, as noted above, the improved conceptual understanding produced through the IE course may have supported the improved physics problem solving. That is, the increases in the FCI associated with the IE course may have been a key factor in mediating the performance gains found on the MB test (reported in Hake [3] and in Crouch and Mazur [11]; but see Thacker *et al.* [12] for dissociations between performance patterns across courses for a conceptual vs a quantitative physics problem). In sum, the interpretation of the few studies that have included quantitative problem-solving outcomes is equivocal.

B. The present experiment and theoretical hypotheses

The present quasiexperiment directly compared learning outcomes from an IE-based course in introductory physics (for convenience we label this the IE course, described in detail in the next section) to a traditional lecture-based course (labeled *traditional*) taught at the same university in the same semester. In contrast to previous studies, our primary focus was on gains on quantitative problem solving, as assessed by a set of 3 typical problems requiring multiple quantitative steps that introductory physics students would be expected to be able to learn to solve in a physics course (see the Appendix). In this regard, an important feature of our design was that we assessed precourse and postcourse performance on these problems. To gauge the extent to which the expected advantages of the IE course on students' conceptual understanding might underlie or relate to the problem-solving outcomes, we also assessed pre- and postperformance on the Force Concept Inventory [14].

Another notable feature of our method was to attempt to externalize and measure various key components of students' problem-solving processes as they attempted to work the problems. Specifically, we required students (i) to bring core concepts of the problem to bear on setting up the problem representation and framing a solution approach (for ease of exposition we will term this *translation*); (ii) to provide the computations to derive an answer (we term this *execution*), and (iii) to evaluate the plausibility of the answer (termed *evaluation*). Additionally, based on a detailed rubric developed by several of the co-authors (detailed in the Sec. II), we awarded partial credit for each of these components. This method allowed us to investigate possible differences across the IE and traditional courses in terms of components of solving quantitative physics problems that might be differentially favored (in terms of student learning) by IE versus traditional courses. Below we detail possible hypotheses in this regard.

One hypothesis is that the IE course will improve performance on quantitative physics problems relative to the traditional course. The reasoning underlying this hypothesis is that increased conceptual understanding promoted by the IE course (as evidenced by an overlapping sample to that of the current experiment [1]) and the interactive problem solving in the IE classroom will support a more "expertlike" approach to physics problem solving [8,15,16]. Findings showing IE-related advantages (relative to traditional lecture courses) on both the FCI and the mechanics problem solving assessment (MB) suggest the plausibility of this hypothesis [3]. Further, we expected that the more expertlike physics problem solving hypothesized to result from the IE course (relative to the traditional course) would emerge particularly in the translation component of problem solving. The idea here is that the increased conceptual understanding promoted by the IE course would presumably facilitate extracting the underlying conceptual principles of the problems [8] and

facilitate the creation of a schema to guide problem solution [15].

An alternative hypothesis is that IE will increase conceptual understanding (as indexed by the FCI) but not necessarily increase proficiency on solving quantitative physics problems. This hypothesis is based on the observation that IE physics courses increase class activities on developing student understanding of fundamental concepts and principles, partly through engagement with conceptual problems, but do not provide more practice or emphasis on quantitative problem solving relative to a traditional physics course (at least, this is the case for the IE course examined in the current study). It is important to note that the IE course in the current study did not focus on conceptual problems at the expense of time spent on quantitative problem solving in class and in homework assignments (but there was less focus on derivations; see Sec. II for details). The underlying assumption of the present hypothesis is that a focus on conceptual aspects does not support transfer to quantitative problem solving (even when there is approximately equal focus in IE and traditional physics on quantitative problem solving). This outcome would be in line with instructors' aspirations that IE can increase conceptual understanding with no penalty to quantitative problem solving relative to traditional courses [3]. These expectations have been expressed by instructors of introductory physics who have adopted IE approaches in their courses (e.g., P. Gibbons, private communication; J. Mestre, private communication), and somewhat supported by one previous study [13].

In line with the above hypothesis, we thought it possible that the particular pattern of quantitative problem solving outcomes from the IE and traditional courses might reflect the basic cognitive science findings regarding problem solving. The basic problem solving research suggests that transfer from practice problems (as given in class and for homework) to new (target) problems (as used on exams and assessments) is greatly influenced by the overlap in features between practice problems and target problems [17,18]. The basic finding is that the more overlap between practice problems and target problems, the better performance will be on the target problems. This finding converges with the seminal "common elements" transfer theory of Thorndyke and Woodworth [19], and with the more contemporary transfer-appropriate processing approach to learning and transfer [20,21,22]. The basic idea is that learning will lead to transfer to the extent that the elements focused on in learning are those reflected in the target task (problem).

This theoretical reasoning leads to the prediction that IE will lead to better quantitative problem solving than the traditional course for target problems with features that align with those emphasized in the IE course but not the traditional course. For the present IE course these are

problems that are context rich and often characterized by real-world "applications" of physics, such as determining if a person's death from being hit by a concrete cylinder rolling down a hill was accidental or due to criminal intent. By contrast, the traditional course will produce better quantitative problem-solving performance than will the IE course for problems presented in typical physics scenarios (not context-rich problems). For the traditional course, these tended to be problems that require applying a sequence of formulas and numerous computational steps (in contrast to the quantitative problem in Thacker *et al.* [12]). These problems require the student to determine the sequence of computational steps and the equations necessary to solve the problem. For a problem not biased toward either set of features (i.e., incorporating a mix of those features), the expectation is that the courses would not be significantly different in performance on a quantitative physics problem. Note that the essence of this set of predictions is that effects of one course format (e.g., IE) relative to the other (e.g., traditional) on quantitative problem solving will follow principles of problem solving and transfer illuminated in the basic learning literature—students will show better transfer to problems of the type on which the instruction focuses. On this view, it is not the case that one course format necessarily generally advantages nor generally penalizes solving quantitative physics problems.

To provide sensitivity to this possibility, the three problems selected for the assessment reflected a differential mix of features just mentioned. One was a context-rich problem presented as an application of a classic physics problem to a "real-world" scenario, requiring students to formulate a problem-solving framework after identifying key physics concepts and parameters of interest, to execute the necessary mathematical calculations, and to draw an inferential conclusion (was a death accidental or intentional; see problem 1 in the Appendix); this problem better aligned with how problems were presented in the current IE course than in traditional physics. Another problem was a classic mathematical physics problem (problem 3 in the Appendix) in which students had to identify and calculate parameters of interest through multiple steps but were provided with the key physics concepts at play. This problem is better aligned with the type of problems solved in traditional physics than in the IE physics course examined here. A third problem (problem 2 in the Appendix) reflected a mix of these features: The physics is wrapped in a real-world context (context rich) and the student needs to identify the key physics concepts involved in order to develop a problem-solving framework (like the IE class problems), but the solver is directed to calculate something specific and is not required to draw a conclusion from the calculations (like the traditional class problems but unlike problem 1 in the Appendix—see part b to problem 2 vs problem 1).

II. METHOD

A. Sample

A total of 194 students enrolled in traditional lecture sections of the 2009 fall semester of introductory physics, and 293 students enrolled in 2009 in parallel IE-based sections participated in the study (these curricula are described below). Not every student in the course participated in the study, but the rate of participation was quite high: 87% and 94% participation rate from the traditional and IE sections, respectively. For traditional physics, two sections were taught, each by a different instructor (total class size was 100 in one section and 122 in the other section). For IE physics, three sections were taught, with two sections taught by the same instructor (total class sizes of 106 and 99) and one section taught by a different instructor (total class size of 107).

Of the 194 traditional-physics participants, 181 students provided gender information; 34% were female and 66% were male. Of the 293 IE participants, 285 provided gender information; 33% were female and 67% were male. For the traditional physics sections, 156 participants provided ethnicity data; 52% were white, 37% were Asian, and 11% were underrepresented minorities (URM). For IE physics, 252 provided ethnicity; 59% were white, 32% were Asian, and 9% were URM. Thus the gender and the ethnicity distributions for the traditional and IE sections were virtually identical. We also were able to determine the majors for 183 students from the traditional sections and 272 students from the IE sections. Both the traditional and IE sections were about evenly divided between engineering majors (49% for each section) and arts and sciences majors (47% for traditional and 46% for IE). Within arts and sciences majors, for traditional physics, half were STEM majors and half were non-STEM majors (just under 25% of total course enrollment for each); for IE physics, 2/3 were STEM majors (30% of total enrollment) and 1/3 were non-STEM majors (16% of total enrollment).

In the traditional physics sample, one student did not provide a complete protocol for the problem-solving assessment (both pre and post-tests); in the IE physics two students did not provide a complete protocol. That is, these students failed to even attempt a solution for at least one problem on the post-test. These 3 students were not included in the overall analyses, but they were included for separate analyses of each problem.

B. Course curricula

1. Traditional-lecture physics

The textbook was *University Physics with Modern Physics* [23]. As detailed in Cahill *et al.* [1], class time was spent primarily covering material in a lecture style, with intermittent demonstrations. The lecture content included working through derivations and problems on the board. When working through problems, instructors

modeled and encouraged use of diagrams as an initial component of solving problems. Evaluation of the plausibility of answers was also modeled but there was variation across instructors and particular problems in the extent to which this occurred. Students enrolled in traditional physics had weekly homework and three noncumulative exams. In addition, extra practice problems were suggested, but these were not required. The homework and extra-practice problems were both end-of-chapter textbook problems and problems written by the instructors; these problems were relatively complicated multistep problems and were assigned after the pertinent lecture(s).

2. Interactive-engagement physics

The textbook was *Six Ideas That Shaped Physics* [24]. This course was oriented toward developing a deeper conceptual understanding of fundamental physics principles and constructs, as opposed to rote learning, as well as to develop quantitative problem solving skills. In order to achieve this, the primary interactive engagement strategies used were demonstrations (including making predictions), two-minute problems, and group work. The two-minute problems are multiple-choice or true-false problems that focus on misconceptions and on conceptual ideas in the chapter. Students discussed the problems in small groups and then voted on the answer. Groups then volunteered to explain their reasoning to the class as a whole. Additional student discussion or instructor feedback or both could be used to orient students toward the correct answer. The majority of the two-minute problems were taken from the textbook or Mazur's ConcepTests [7]; some were constructed by the IE course instructors. Depending on instructor, the interactive engagement activities represented 40%–60% of class time, and could include some quantitative problem solving. The remainder of the class time was spent on “mini lectures” that could include instructor-led examples (working examples) and derivations. As in the traditional physics course, when working example problems, instructors in IE physics modeled and encouraged the use of diagrams (generated and labeled diagrams); there was variation across instructors and problems in the extent to which evaluating plausibility of answers was modeled. Essentially, in terms of class time, the IE physics course examined in the current study spent time on solving and discussing conceptual problems but not at the expense of modeling and working through quantitative problems. Instead, in the IE physics course there was less frequent and less detailed development of derivations (relative to the traditional physics course examined here).

Both traditional and IE physics used demonstrations in their classes but the way students interacted with those demonstrations was different. In traditional physics the demonstration was shown and then the instructor proceeded to solve different problems based on the concepts that were shown in the demonstration. By contrast, in IE

physics, before the demonstration was performed students generally divided into groups, the groups tried to predict the outcome of the demonstration, and these predictions were shared and defended with the class. After these interactions, the demonstration was performed. The results of the demonstration were then explained either by students in small groups, by the instructor, through a series of two-minute problems, or most often by some combination of these methods. Some time was spent on derivations in the IE physics sections, but derivations were given less weight than in typical lecture classes.

Students enrolled in IE physics had daily homework covering the daily reading and weekly homework covering material that was both conceptual and quantitative in nature. The assigned weekly homework load was fairly comparable to the weekly homework in the traditional-physics classes, with 3 daily homework assignments (2 problems each) and 1 weekly assignment (3 problems); in traditional physics 5–8 homework problems were assigned weekly and another 5–8 problems were suggested as practice. No additional problems were explicitly suggested for practice in the IE course. For both courses, the homework was graded in terms of several components, including translation (diagrams were required) and evaluation of plausibility of answers. The main differences were (i) that unlike homework in traditional physics, the IE physics daily homework included conceptual problems (not purely quantitative problems), (ii) because daily IE-physics homework assignments were completed before class, the quantitative problems were by necessity not as advanced or multistep (as in traditional physics homework), and (iii) following the textbook’s suggestion [24], to encourage students in IE physics to learn from their errors, students were given the opportunity to revise their homework to earn additional points. Answers were provided online, and students could try to figure out why points were lost as they revised their homework, thereby providing an additional learning opportunity.

Students in IE physics took three noncumulative exams, as did students in traditional physics. The exams for IE and traditional physics were similar in that they consisted primarily, if not entirely, of quantitative problems requiring multiple steps. Typically the exams included 3–6 problems, depending on the number of parts in each problem. One difference in the exams (across IE and traditional physics) was in the framing (or style) of the problems. The IE physics exams tended to be set up in a context-rich framing (e.g., see problem 1 in the Appendix), whereas the traditional physics exams used textbook-style problems. Second, the presentation of the IE exam problems often specified the type of problem (e.g., “conservation of momentum”; “use energy”) and asked students to generate diagrams and label picture components as a first step in solving the problem; this component was graded. By contrast, in the traditional-course exams the problem was simply presented without specification of the problem

type and scoring focused on the correctness of the answer (though partial credit for appropriate translation could be earned when the problem answer was incorrect).

C. Assessments

To assess conceptual learning, following previous studies [1], we used a physics concept inventory appropriate for the content of the initial semester of introductory physics—the Force Concept Inventory (FCI) [14]. The FCI is a 30-item multiple-choice test that assesses basic knowledge of mechanics concepts covered in the physics course.

The problem-solving assessment contained three problems constructed by one of the course instructors (who taught IE sections). Problem 1 involved conservation of energy using rotational kinetic energy in a real-life application. Problem 2 involved conservation of energy taking into account thermal energy. Problem 3 was a classic conservation of energy problem involving friction. The three problems are presented in the Appendix. For each of the three problems, students were asked three questions designed to elicit (a) the key concepts, principles, assumptions, and plan for solving the problem, (b) the execution of the plan and reasoning in terms of the computational work (see the Appendix), and (c) evaluation of the plausibility of the answer (see the Appendix for the wording of these questions).

The FCI and the problem solving assessment were administered both at the outset of the semester (“precourse”) and near the end of the semester (“postcourse”). Details of the administration of these assessments are provided in Sec. II E below.

D. Problem scoring

A scoring rubric for each component of a problem (translation, execution, and evaluation) was initially created by an instructor from the course and substantially refined and detailed by one of us. This detailed rubric was then refined through discussion with two of the other authors. A brief description of the scoring approach is provided next (the fully detailed scoring rubric is available on request from the authors). The rubric was constructed to directly evaluate the problem-solving components elicited by the current problem-solving assessment (i.e., questions a, b, and c for each problem); however, the rubric captures aspects of physics problem solving reflected in a recently published and validated rubric applied to introductory physics problem solving (e.g., scores reflect “a useful description” and the “physics approach,” and “mathematical procedures” [25]).

1. Translation

There were two aspects of translation extracted in the scoring, and for the present analyses these were blended

into a single score that was included in a total problem-solving score (described below). The first aspect captured how effectively the student interpreted the problem and developed and communicated a model that could be used to solve the problem. The possible scores ranged between 0 and 3. A “0” indicated a complete failure to interpret the problem correctly, a “1” indicated a major misinterpretation of the problem; a “2” indicated a minor misinterpretation of the problem, and a “3” indicated that a student correctly interpreted the problem, which involved creating a model for the situation (e.g., drawing a diagram, though the model could also be symbolically or verbally described) that was accurate and logically complete to tackle the problem.

Though providing a diagram was not explicitly required in the problem-solving instructions given to students, many (but not all) students generated a diagram in the translation part of the problem. (Throughout the semester, use of diagrams was encouraged to help students generate an approach for solving any particular problem.) To directly capture the quality of the diagrams that were provided, the second aspect of translation was scoring the diagrams on a scale of 0 to 3. A “0” was assigned when no diagram was generated, a “1” reflected major misinterpretations, a “2” reflected minor misinterpretations, and a “3” was for a complete and fully accurate diagram. The translation score used in the current analyses was the average of the first and second scores just described (separating these two aspects did not change the patterns of results).¹

2. Execution accuracy

This score measured how completely and accurately a student solved the problem or implemented their plan of attack, using valid physics principles and mathematical operations. A student could receive a score between 0 and 3. A “0” indicated an inappropriate or incomplete solution or did not arrive at a final solution, a “1” indicated major mistakes in reasoning or application of physics principles, a “2” indicated minor mistakes in reasoning or application of physics principles, and a “3” indicated the solution was completely correct and appropriate.

3. Evaluation

This score gauged how effectively a student evaluated the magnitude, units, and signs—that is, determined (or

¹We elected to report this particular score because the literature in both physics and chemistry education suggests that drawing diagrams can be an important factor in supporting effective solving of physics [16] (and chemistry [26]) problems, and doing so was emphasized in both the traditional and IE curricula. Accordingly, though the instructions on the problem-solving assessment did not require students to draw a diagram to solve the problems, full credit on the translation component was awarded only if a diagram of some sort was present. In order to not overweight the diagram score, an average of translation and diagram was used for the translation score.

appraised) the plausibility of their results. A student could receive a score between 0 and 3. A “0” indicated no meaningful evaluation or no solution had been obtained; a “1” indicated an evaluation was present but elements or reasoning was missing or incorrect; a “2” indicated a well thought out, meaningful, and accurate reflection on one of the three plausibility features; a “3” indicated a well thought out, meaningful, and accurate reflection on at least two of the three plausibility features.

4. Total problem solving score

To provide an overall summary of problem solving performance, we combined the above subcategories into one total score. We developed four different schemes to create a total score that ranged in value from 0 to 9. For the purposes of this paper, we focus on only one of these schemes because all 4 totals showed the same results.² The total used in the reported analyses was the sum of the translation, execution accuracy, and evaluation scores. The total has possible scores ranging from 0 to 9.

5. Planning coherence

Finally, though some responses did not include an explicitly stated plan for solving the problem, we were interested in assessing the quality of plans that were provided. Accordingly, we applied the following scoring scheme for this problem solving component: A “0” indicated that the plan of attack was missing or meaningless; a “1” indicated that the plan of attack had major inconsistencies; a “2” indicated that the plan of attack had minor inconsistencies; a “3” indicated that a student described a logical plan of attacking the problem (i.e., a plan that is self-consistent and is also consistent with execution and should yield an appropriate, but not necessarily correct, solution) based on a clearly presented reasoning that refers to the specific concepts and equations to be used. We analyze this component separately from the others (see Sec. III).

6. Scoring reliability

Initially one of us scored a set of 89 protocols, with each set reflecting the precourse and postcourse problem-solving responses of a student. Another author independently scored a subset of 30. The two graders were able to apply the rubric with extremely high intergrader consistency (the rubric was appropriately detailed, which likely contributed to the high scoring reliability; cf. [27]). The scores from the two raters were identical, except for diverging by 1 point on one problem in the evaluation category (99% perfect overlap). Because the interrater consistency was

²While the analyses for these totals are not included in the paper, the full analyses are available upon request from the authors.

virtually perfect, the remaining 398 sets of problem-solving responses were scored by one person. For all of the protocols, the graders were blinded to the students' names, gender, lab section, and course (i.e., the protocols were prepared so that all student and course information were removed).

E. Procedure

Students completed the problem solving assessments during their 2.5 h weekly lab sections under the supervision of a graduate teaching assistant. Participation was strongly encouraged, but was not required and incentives, such as extra credit, were not given for completing the assessments. The precourse assessment (we use this term for convenience; the assessment was administered after the course began) was given during the first lab of the semester and students were given all but roughly 20 min of lab period to complete the questions. For most students, this lab occurred during the third week of classes, before the concept of energy was introduced. However, three lab sections of the IE course met after the first completed homework and lecture on energy conservation (16–17 students per lab; 49 total students) and an additional four lab sections met after two days of homework and lecture on energy conservation (64 total students). Because these lectures only addressed the conversion of gravitational potential energy to kinetic energy, the information provided in class and through the homework was insufficient to solve any of the problems on the assessment. The topics of rotational energy, thermal energy, and work were all introduced after the precourse assessments were completed. All of the students in traditional physics completed the precourse assessment before the topic of energy was introduced. Students completed the postcourse assessments during lab 1–2 weeks after they were tested on energy (week 11 of the 14 weeks of classes) and had 1.5–2 h to complete the assessment. The FCI precourse test was completed during the second and third weeks of the semester and the FCI postcourse test was completed 2–3 weeks before the semester ended [1].

III. RESULTS

Initial inspection of the precourse baselines indicated that the baselines for IE physics were higher than for the traditional physics sections. We confirmed this with an analysis of variance (ANOVA) contrasting the two sections: for the FCI, $F(1,318) = 9.34$, mean-squared error (MSE) = 0.046, $p = 0.002$, and for the quantitative problem solving assessment, $F(1,482) = 43.247$, MSE = 8.692, $p < 0.0005$ (the advantage of IE students relative to traditional-physics students on the problem-solving precourse baseline held for all three problems). To take into account these precourse differences in the two instructional conditions, we adopted two standard analytic

approaches. In one approach, we examined the gains from pre- to postcourse for the total sample, doing so with analyses of variance (ANOVA); these analyses are reported first (we also include a normalized-gain analysis). In the other approach, we matched the precourse quantitative problem-solving baselines across the two instructional conditions by using a reduced sample from the IE physics condition. To further equate the samples, we statistically equated pre-FCI scores by conducting analyses of covariance (ANCOVAs) on the reduced sample (with pre-FCI as the covariate); these analyses are reported second. For both approaches we report analyses of the total problem solving score, followed by analyses of each of the components.

A. Full sample analyses

1. Total problem-solving score

We conducted a $2 \times 3 \times 2$ mixed-model ANOVA on the total scores, with instruction type (traditional vs IE) as a between-subjects variable, and problem number and the timing of the test (pre vs post) as the within-subjects variables. As expected, there was a significant increase in performance from pre- to postcourse, $F(1,482) = 319.116$, MSE = 3.94, $p = 0.000$, $\eta^2 = 0.398$. This performance increase was slightly but significantly more pronounced for problem 3, $F(2,964) = 5.017$, MSE = 2.68, $p = 0.007$, $\eta^2 = 0.010$, for the interaction. Overall (collapsed across pre- and postcourse), the problems significantly differed in how well they were solved, $F(2,964) = 45.489$, MSE = 3.98, $p = 0.000$, $\eta^2 = 0.086$, with problem 1 solved the best and problem 3 solved the least well (Fig. 1 displays the means). Though this effect could reflect problem difficulty, it may also be that students were more fatigued when solving the last problem than when solving the first problem.

Regarding the effects involving instruction type, overall (pre- and postcourse test combined), students in IE physics performed significantly better than students in traditional physics, $F(1,482) = 23.532$, MSE = 14.53, $p = 0.000$,

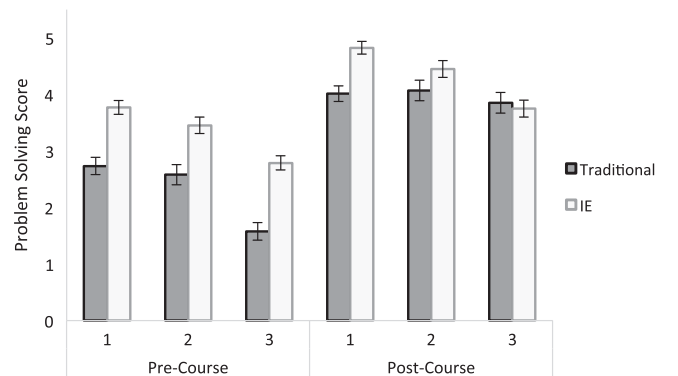


FIG. 1. Pre- and postcourse total problem-solving score as a function of problem number and class format (total possible score = 9). Error bars represent the standard error.

$\eta^2 = 0.047$. More critically, for the central question of interest, the magnitude of gain from pre- to postcourse test differed significantly as a function of instruction type, $F(1, 482) = 20.260$, $MSE = 3.94$, $p = 0.000$, $\eta^2 = 0.040$ (for the two-way interaction). Inspection of Fig. 1 reveals that the increase in problem-solving scores was more robust for traditional physics than for IE physics instruction. This pattern further interacted with problem number, $F(2, 964) = 6.971$, $MSE = 2.68$, $p = 0.001$, $\eta^2 = 0.014$, such that the advantage for traditional physics in problem solving gains was apparent for problems 2 and 3 but not problem 1. To confirm this interpretation, for each problem we conducted a 2 (instruction type) \times 2 (pre, post) mixed ANOVA. For problems 2 and 3, traditional physics produced significantly larger gains from pre- to postcourse than did IE physics, $F(1, 484) = 4.05$, $MSE = 3.71$, $p = 0.045$, $\eta^2 = 0.008$; and $F(1, 483) = 37.068$, $MSE = 2.69$, $p < .001$, $\eta^2 = 0.07$, respectively. By contrast, for problem 1 the gains were not significantly different across traditional and IE physics ($F < 1$, $p = 0.318$, $\eta^2 = 0.002$).

We also analyzed normalized gains (as is customary for concept inventory scores), using the normalized change approach specified in Marx and Cummings [28]. A 2 (instruction type) \times 3 (problem number) mixed-model ANOVA yielded results that were similar to those just reported. In general, normalized gains were significantly larger for traditional physics ($M = 0.22$) than IE physics ($M = 0.17$), $F(1, 472) = 6.308$, $MSE = 0.170$, $p = 0.012$, $\eta^2 = 0.013$. Further, there was a significant interaction with problem number, $F(2, 944) = 6.636$, $MSE = 0.113$, $p = 0.001$, $\eta^2 = 0.014$, such that for problems 2 and 3 gains were more prominent for traditional physics [Means (Ms) = 0.21 and 0.29, respectively] than for IE physics (Ms = 0.17 and 0.15, respectively); for problem 1 gains were comparable (Ms = 0.17 and 0.18 for traditional and IE physics, respectively).

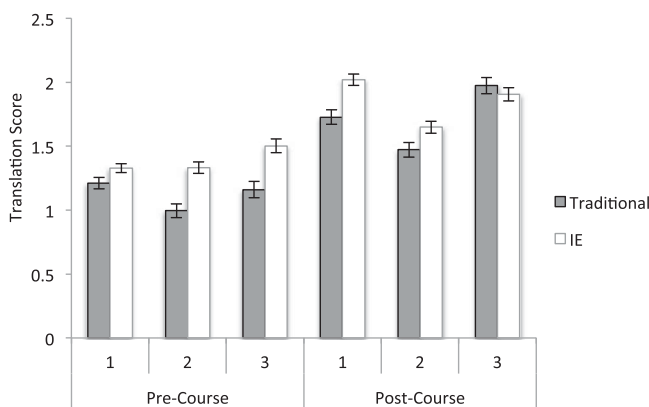


FIG. 2. Pre- and postcourse translation score as a function of problem number and class format (total possible score = 3). Error bars represent the standard error.

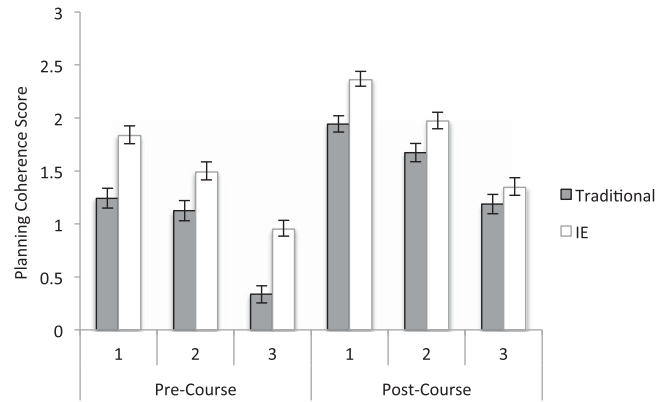


FIG. 3. Pre- and postcourse planning coherence score as a function of problem number and class format (total possible score = 3). Error bars represent the standard error.

2. Problem-solving components

To determine the components of problem solving that were advantaged by the traditional physics, and also whether the initial translation of the problem (translation) and explication of a solution plan (planning coherence) might have been better for IE physics, for each problem we conducted 2 (instruction type) \times 2 (pre vs post) ANOVAs that examined each component of the problem solution (translation, planning coherence, execution accuracy, evaluation). For all components and all problems there were significant gains from the pre- to postcourse test (for purposes of efficiency the F values are not reported). We focus here on the interactions between instruction type and pre-post course gains. Paralleling the overall problem-solving score, for the translation component the gains from pre- to postcourse were more robust for traditional physics than for IE physics for problems 2 and 3, $F(1, 484) = 3.31$, $MSE = 0.446$, $p < .07$, $\eta^2 = 0.007$ and $F(1, 483) = 21.375$, $MSE = 0.463$, $p < 0.0005$,

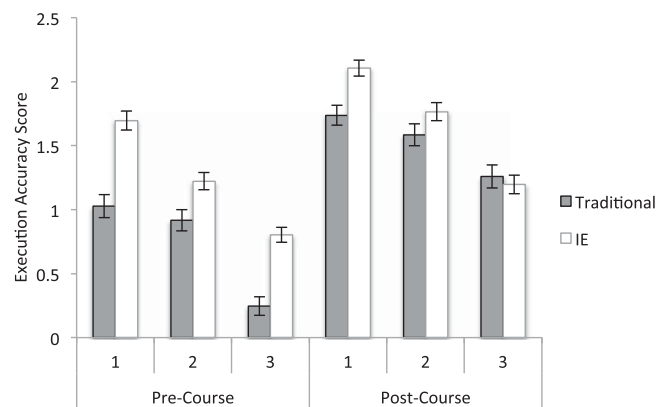


FIG. 4. Pre- and postcourse execution accuracy score as a function of problem number and class format (total possible score = 3). Error bars represent the standard error.

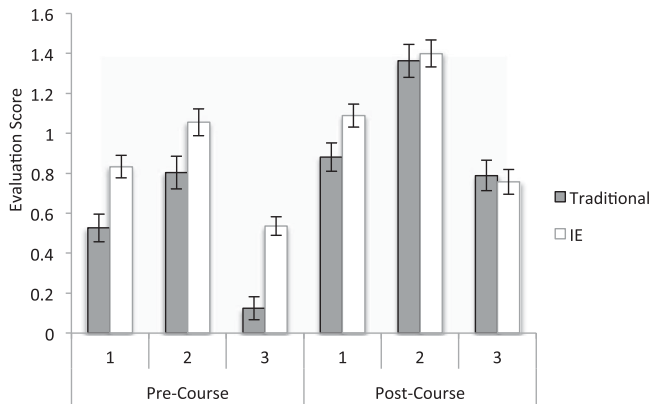


FIG. 5. Pre- and postcourse evaluation score as a function of problem number and class format (total possible score = 3). Error bars represent the standard error.

$\eta^2 = 0.042$, respectively (see Fig. 2). As anticipated, for problem 1 IE physics produced more gain in the translation score than did traditional physics, $F(1, 485) = 4.335$, $MSE = 0.424$, $p = 0.038$, $\eta^2 = 0.009$. For plan coherence, traditional physics produced more substantial gains than did IE physics for problem 3, $F(1, 483) = 17.350$, $MSE = 0.706$, $p < 0.0005$, $\eta^2 = 0.035$; no significant interactions emerged for problems 1 and 2 (see Fig. 3).

For execution accuracy, traditional physics produced significantly more gain in accuracy from pre- to post-test for problems 1 and 3 than did IE physics, $F(1, 485) = 5.457$, $MSE = 0.974$, $p = 0.020$, $\eta^2 = 0.011$ and $F(1, 483) = 32.235$, $MSE = 0.691$, $p < 0.0005$, $\eta^2 = 0.063$, respectively (see Fig. 4). There was no interaction for problem 2. For evaluation, traditional physics also produced more robust gains than IE physics for problem 3, $F(1, 483) = 25.239$, $MSE = 0.457$, $p < 0.0005$, $\eta^2 = 0.050$; no interactions emerged for problems 1 and 2 (see Fig. 5). On balance then, the component of problem solving most consistently advantaged by traditional physics was execution accuracy. In addition, traditional physics favored substantial gains on problem 3 for all problem solving components, a pattern we return to in Sec. IV.

3. FCI score

We conducted a 2 (instruction type) \times 2 (pre vs post) mixed ANOVA on the FCI scores that were available. (Because the focus in this study was on quantitative problem solving, less emphasis was placed on students completing the FCI, and thus a number of students completed the post problem-solving assessment but did not complete the post FCI and could not be included in the ANOVA. Also these FCI data represent a subset of students that were included in a larger published study that focused on FCI outcomes across three years of IE and traditional physics [1]). There was a significant increase in FCI

performance from the pre- to postcourse test, $F(1, 318) = 153.948$, $MSE = 0.008$, $p = 0.000$, $\eta^2 = 0.326$, and overall (pre and post combined) FCI scores were higher for students in IE physics than in traditional physics, $F(1, 318) = 17.539$, $MSE = 0.072$, $p = 0.000$, $\eta^2 = 0.052$. Most importantly, there was a significant interaction $F(1, 318) = 4.821$, $MSE = 0.008$, $p = 0.029$, $\eta^2 = 0.015$, such that IE physics produced significantly higher gains from the pre- ($M = 0.699$) to postcourse test ($M = 0.804$) than did traditional physics ($M = 0.624$ and 0.697 , respectively). In line with previous studies, we also analyzed normalized gains [1]. Paralleling the just reported interaction, IE physics produced significantly higher normalized gain than did traditional physics, $F(1, 318) = 9.19$, $MSE = 0.098$, $p = 0.0026$.

In sum, as in previous studies [1,3,13], IE physics produced superior gains on concept inventory performance (FCI) relative to traditional physics. The key finding is that the reverse pattern emerged for performance on quantitative physics problems. Here, the traditional physics instruction supported generally higher gains than did the IE physics instruction. Thus, for the two particular instructional implementations examined, conceptual gains (as indexed by the FCI) were dissociated from quantitative problem-solving gains (as indexed by the three-problem assessment). To further explore the relation between the postcourse conceptual knowledge (FCI) and quantitative physics problem-solving performances, we combined the instructional groups to compute the correlation between the post-FCI scores and the postcourse scores for each of the quantitative problems. The correlations between the postcourse FCI and problem-solving total scores were significant but modest, ranging from $r(319) = 0.268$, $p < 0.0001$, for problem 2 to $r(319) = 0.307$, $p < 0.0001$, for problem 3. That is, less than 10% of the variance in quantitative problem solving performance was related to the end-of-course FCI scores. This pattern reinforces the above observation from the experimental

TABLE I. Total precourse mean problem-solving scores (and standard deviations) and increase in postcourse total problem-solving scores for sample of IE physics students matched to the traditional physics students on the precourse scores. (Δ Mean indicates the average of the difference between each student's pre and postcourse score.)

Problem number	Course format	Precourse mean	Std. dev	Δ Mean	N
1	Traditional	2.744	1.791	1.270	176
	IE	2.696	1.791	1.841	202
2	Traditional	2.489	2.064	1.602	175
	IE	2.537	2.067	1.487	231
3	Traditional	1.548	1.340	2.338	176
	IE	1.554	1.340	1.375	212

results that the standard physics concept inventory used in many studies to index gains in conceptual knowledge from introductory physics courses is not necessarily tightly intertwined with the learning of quantitative problem solving, at least for beginning physics students.

B. Matched sample analyses

1. Total problem-solving score

For each problem, we excluded the higher scoring students from IE physics such that we configured a sample for which the mean pretest total score was as close to the mean for the traditional physics students as possible (we did this for each problem separately, to maximize the comparability of the matched samples). The precourse scores and standard deviations, as well as the postcourse gains are summarized in Table I for each problem. To further equate the precourse physics knowledge for students in IE and traditional physics, we used the students' precourse FCI scores as a covariate in the analysis. Thus, for each problem we conducted a mixed analysis of covariance (ANCOVA) with instruction type as the between-subjects variable and test time (pre vs post) as the within-subjects variable.

TABLE II. Precourse translation scores (and standard deviations) and increase in postcourse translation scores for sample of IE physics students matched to the traditional physics students on the precourse translation scores. (Δ Mean indicates the average of the difference between each student's pre and postcourse score.)

Problem number	Course format	Precourse mean	Std. dev	Δ Mean	N
1	Traditional	1.205	0.570	0.511	176
	IE	1.218	0.569	0.757	202
2	Traditional	0.960	0.635	0.514	175
	IE	1.095	0.638	0.446	231
3	Traditional	1.114	0.783	0.875	176
	IE	1.170	0.786	0.500	212

TABLE III. Precourse coherence scores (and standard deviations) and increase in postcourse coherence scores for sample of IE physics students matched to the traditional physics students on the precourse coherence scores. (Δ Mean indicates the average of the difference between each student's pre- and postcourse score.)

Problem number	Course format	Precourse mean	Std. dev	Δ Mean	N
1	Traditional	1.239	1.327	0.744	176
	IE	1.426	1.322	0.886	202
2	Traditional	1.069	1.270	0.657	175
	IE	1.177	1.277	0.641	231
3	Traditional	0.358	0.915	0.864	176
	IE	0.410	0.917	0.552	212

TABLE IV. Precourse execution accuracy scores (and standard deviations) and increase in postcourse execution accuracy scores for sample of IE physics students matched to the traditional physics students on the precourse accuracy scores. (Δ Mean indicates the average of the difference between each student's pre and postcourse score.)

Problem number	Course format	Precourse mean	Std. dev	Δ Mean	N
1	Traditional	1.028	1.207	0.728	176
	IE	1.178	1.208	0.782	202
2	Traditional	0.863	1.045	0.748	175
	IE	0.866	1.049	0.731	231
3	Traditional	0.256	0.650	1.022	176
	IE	0.245	0.655	0.595	212

TABLE V. Precourse evaluation scores (and standard deviations) and increase in postcourse evaluation scores for sample of IE physics students matched to the traditional physics students on the precourse evaluation scores. (Δ Mean indicates the average of the difference between each student's pre- and postcourse score.)

Problem number	Course format	Precourse mean	Std. dev	Δ Mean	N
1	Traditional	0.540	0.809	0.358	176
	IE	0.401	0.796	0.619	202
2	Traditional	0.794	1.005	0.555	175
	IE	0.671	1.201	0.554	231
3	Traditional	0.136	0.464	0.659	176
	IE	0.113	0.466	0.340	212

For problem 3 the results were similar to those from the ANOVA conducted on the overall sample. Specifically, the increase in the problem solving score from pre- to postcourse test was significantly more robust for traditional physics than for IE physics instruction, $F(1, 385) = 19.730$, $MSE = 2.54$, $p < 0.0005$, $\eta^2 = 0.049$. By contrast, for problem 2 the gains in the problem solving for traditional and IE physics did not significantly differ ($F < 1$). The most prominent difference from the overall ANOVA was for problem 1; gains in problem solving were significantly more substantial for IE physics than for traditional physics, $F(1, 375) = 6.666$, $MSE = 2.55$, $p = 0.010$, $\eta^2 = 0.017$.

2. Problem-solving components

As for the full-sample analyses, for the matched samples we examined performance on each problem-solving component for each problem using ANCOVAs with pre-FCI as the covariate. Tables II–V display the precourse scores and standard deviations, and the gains at postcourse for the translation, planning coherence, execution accuracy, and

evaluation scores, respectively. For problems 2 and 3, for every component the patterns paralleled that found with the total problem solving score. Thus for problem 2, traditional and IE physics produced equivalent gains from pre- to post-test for every component (all F s < 1). For problem 3, traditional physics produced significantly larger gains than did IE physics for every component: Translation, $F(1, 385) = 14.373$, $MSE = 0.47$, $p < 0.005$, $\eta^2 = 0.036$; planning coherence, $F(1, 385) = 7.327$, $MSE = 0.73$, $p = 0.007$, $\eta^2 = 0.019$; execution accuracy, $F(1, 385) = 16.092$, $MSE = 0.64$, $p < 0.0005$, $\eta^2 = 0.040$; evaluation, $F(1, 385) = 14.348$, $MSE = 0.40$, $p < 0.0005$, $\eta^2 = 0.036$. For problem 1, the greater overall problem solving gains for IE relative to traditional physics was primarily a consequence of significantly more robust gains (for IE) for translation of the problem, $F(1, 375) = 6.838$, $MSE = 0.40$, $p = 0.009$, $\eta^2 = 0.018$, and evaluating the plausibility of the computed solution, $F(1, 375) = 5.133$, $MSE = 0.611$, $p = 0.024$, $\eta^2 = 0.014$. The pre- to post-test gains in coherence and execution accuracy did not significantly differ for IE and traditional physics (largest $F = 1.03$, for coherence).

IV. DISCUSSION

Two central findings emerged from this study. First, the IE physics course that was implemented produced greater gains in conceptual understanding of physics, as indexed by a widely used physics concept inventory (the FCI), than did the traditional physics course. Second, and more novel, the conceptual gains observed for the IE physics course were not paralleled by the quantitative problem-solving outcomes. Across the entire sample of students, the traditional physics course produced greater gains in quantitative problem solving than did the IE physics; with the reduced matched sample, IE physics still did not produce overall advantages in quantitative problem solving relative to traditional physics. We discuss the FCI results first before turning to the more provocative and novel quantitative problem-solving results.

The positive effect of IE relative to traditional physics for improving conceptual understanding in physics is itself an important addition to the research literature examining the impact of interactive-engagement techniques in general physics courses (and science courses more generally). Though a fairly substantial literature has reported that IE innovations generally are associated with better postcourse performance (measured through exam performances and concept inventories) [3,4], the interpretation of the majority of these results (and associated meta-analyses) is seriously compromised by a major shortcoming. Specifically, across 310 studies (contrasting students given IE enriched instruction versus standard instruction in biology, chemistry, engineering, or physics) published in peer-reviewed journals from 1990 to 2007, of the 89% that were

quasiexperiments (students were not randomly assigned to instructional condition), two-thirds failed to administer pretests, “making it impossible to rule out preexisting differences in achievement between groups that could artificially inflate the effect of innovations” [5]. The real possibility of preexisting achievement differences between groups (for the quasiexperimental designs that are overwhelmingly implemented in this literature) is underscored by the present findings showing that the students enrolled in the IE physics course had significantly more knowledge in physics as a group (in terms of both conceptual understanding and quantitative physics problem solving) than did the students in the traditional physics classes at the outset of the course. Clearly, to the extent that this difference has existed in past studies reporting positive impact of IE courses on postcourse performances, previous research-based claims favoring IE over standard physics courses need to be cautiously interpreted.

The present results focusing on gains from pre- to post-test FCI performances thus provide important convergence to the minority of published studies (that have considered pretest achievement) showing positive impact of IE techniques on increasing conceptual understanding in physics relative to the traditional lecture [1,2,13]. Even taking pretest performance into consideration, we hasten to add that in a quasiexperimental design (like that used here and in most previous studies) significant differences in the proportions of males and females, minority status, or college major between the two course formats could potentially impact outcomes. Fortunately, the present samples were nearly identical in their gender and minority composition for the IE and the traditional physics classes, as were the proportions of engineering versus arts and sciences majors. Thus, the present advantage of IE over traditional physics cannot be a consequence of differential distributions of gender, ethnicity, or college majors (i.e., engineering or arts and sciences) across the two courses. Instead, the most likely interpretation of our FCI results is that the instructional approach in IE physics supports more conceptual understanding of the physics principles tested on the FCI than did the traditional physics instruction.

A. Performance on quantitative physics problems

The more novel component of the current study was the focus on learning to solve quantitative physics problems. This is a core objective in most college-level general physics courses, as these courses are in part preparing students for upper-level physics, science, and engineering courses. Positively, regardless of instruction type, on three types of conservation of energy problems students showed significant performance gains in solving these problems from pre- to postcourse. A more central finding for the present purposes was that when all students were considered in the analyses, the magnitude of these gains varied depending on instruction type, favoring the traditional

instructional condition more so than the IE instruction (on two of three problems). It is important to note that these analyses including all students reflect the standard analytic procedure for quasiexperimental studies that collect pre-course baseline data: The gains from pre- to post-test are viewed as taking into account possible precourse differences in student knowledge across the instructional conditions [13]. From this perspective, the above patterns contribute a broader picture of the outcomes of IE and traditional physics instruction than currently provided in the literature. For this pool of students, the present findings show more robust gains on learning conceptual components of physics (i.e., FCI scores) with IE-enriched physics [1,3,13], but more robust gains on learning to solve quantitative physics problems with traditional physics instruction.

These quantitative problem-solving results may appear to contradict previous reports showing advantages on problem solving tasks after IE relative to traditional physics instruction [3,11,12]. Those previous findings are inconclusive, however, as they did not include pretests (see also Ref. [13]); it is possible that gains from pre- to post on the quantitative physics problems in those previous studies would not have been advantaged by IE physics. Also, two of those studies [3,12] used relatively simple quantitative problems (as outlined in the introduction), so even if subsequent work were to verify those patterns (using pretest designs), the current results at the least indicate that for moderately difficult quantitative problems, traditional physics can produce greater learning than IE physics (as implemented in this study).

The above conclusions favoring traditional physics might be somewhat tempered by the observation that students enrolled in IE physics showed significantly higher precourse physics knowledge than did the students in traditional physics on both the concept inventory and quantitative problem-solving assessments. The concern here is that the higher baseline for students in IE physics would reduce the magnitude of gains that could be achieved relative to students in traditional physics. Mitigating this concern are three observations. First, the baseline scores were far from ceiling for the IE physics condition (the total quantitative problem-solving baseline score averaged 3.34 over the three problems, out of a total possible of 9.0). Second and importantly, the normalized gains analysis—standardly used to adjust for possibly limited absolute gains associated with higher baselines—also showed significant gains for traditional physics relative to IE physics on quantitative problem solving. Third, even with higher baseline scores on the FCI, IE physics still produced significantly greater gains than did traditional physics on the FCI. However, the learning trajectory and dynamics for learning how to solve quantitative physics problems might not parallel that for conceptual understanding, especially with the relatively difficult quantitative problems used on

the current assessment. Specifically with regard to the current patterns, gains from higher problem-solving baselines may be more difficult to achieve than gains from lower baselines. Importantly, in line with this possibility, the analyses using the restricted IE physics sample that equated baselines with the traditional physics sample showed a pattern that indicated a less general advantage for traditional physics. Thus, at the very least, our findings strongly support the conclusion that an IE enriched physics course that produced more robust gains relative to the traditional physics course in conceptual understanding (on the FCI) did not produce parallel advantages relative to the traditional course in learning to solve quantitative physics problems.

B. Implications

These results, regardless of whether considering the full sample or the restricted matched sample, do not support the hypothesis (developed in the introduction) that increased conceptual understanding supported by IE physics might also generally improve quantitative problem solving skills taught in the introductory physics course. First, the effects of instruction dissociated across the concept inventory and the quantitative problem solving assessment (such that gains on the conceptual inventory in IE physics were not matched and were even reversed on some of the quantitative problems). Second, the correlation between the postcourse performances on the two assessments was relatively low (with FCI accounting for less than 10% of the variance of the performance on the quantitative problems). This finding is consistent with an existing computer model of introductory-physics students' conceptual knowledge and quantitative-problem knowledge in classical mechanics, which indicated only modest overlap in the two types of knowledge [29].

These patterns suggest that the skills and knowledge required for quantitative problem solving in physics may be somewhat independent of (or more involved than) the conceptual learning supported in IE-enriched physics through working on nonquantitative “conceptual” problems. This suggestion is consistent with literature on expertise indicating that conceptual and problem-solving knowledge might not be integrated until later in the development of expertise (see Refs. [30] and [31] in the domain of physics). Of course, the present study did not closely link the conceptual index (FCI), which emphasized knowledge of Newton's three laws, to the particular quantitative problems included in that assessment (focusing on conservation of energy). It remains possible that if the concepts for which the IE course showed greater gains were embedded in the quantitative problems (see, e.g., Ref. [16]), then IE might have produced similar or increased gains (than traditional instruction) on these problems, especially with concentrated practice integrating

the concepts with the quantitative problem solving [32,33]. More generally, the current problem-solving assessment was limited to work-energy problems,³ and the results from these problems might not have extended to problems sampled from other topics of the introductory course.

Still, the orientation in the present IE-enriched course focused on class problem-solving activities that emphasized approaching quantitative problems by bringing core concepts of the problem to bear on setting up the problem representation and framing a solution approach (termed *translation* in our problem-solving scoring rubric). This also occurred in the traditional physics lectures. However, a major difference was that the exams in the current IE course directed students at the outset of the exam problems to set up a problem representation (generate a diagram or label parts of a diagram) and exam scores reflected grading of this component. In contrast, in traditional physics the exam problems were simply presented and getting the correct answer was the focus of the exam scoring. Accordingly, it might have been expected that, at least on the translation component, that IE-enriched physics would generally support more gains than traditional physics. That clearly was not the case, except for problem 1 in the matched sample analysis.

The attempt to emphasize these problem-solving components likely has merit [15,16,26], however, the present findings underscore that improving components of problem solving involves regular deliberate practice and clear feedback [10,16,34]. In this regard, though homework problems in both IE and Traditional physics required translation, the scoring and feedback on this component was typically somewhat cursory. Accordingly, IE enriched physics (indeed, all instructional approaches) might produce greater gains in translation skills (applying relevant concepts to the problems [8] and constructing appropriate representations [26]) by incorporating increasing practice, feedback, or both on these components [16,34]. Indeed, current efforts in our introductory physics courses are incorporating these modifications.

Another hypothesis regarding outcomes from IE-enriched physics is that the IE curriculum at least does not penalize students in terms of learning how to solve quantitative physics problems. Several informal communications from physics instructors who have

implemented IE techniques in general physics have expressed this possibility (P. Gibbons, personal communication, Feb. 9, 2016; J. Mestre, personal communication, Sept. 28, 2015). The present findings are mixed with regard to this hypothesis regarding the current IE-enriched implementation. Considering the entire sample, the results were counter to this hypothesis, as the IE-enriched course produced significantly diminished gains (on quantitative problem solving) than did the traditional physics course. However, to the extent that the matched sample results are considered more revealing, then on balance the IE-enriched course did not penalize students on quantitative problem solving (see Refs. [35,36] for similar findings).

In terms of the quantitative problem-solving results, the outcomes from each instructional approach appeared to generally reflect what was practiced. In the IE course, as is often the case with this instructional method, the quantitative problems were often context-rich and presented in terms of “real-world” applications, whereas in the traditional course the problems were not contextualized in this way. For the matched sample at least, this kind of contextualized problem showed greater gains after IE than traditional instruction. The IE course did include class instruction on standard multistep quantitative physics problems, but the traditional course required more practice on these types of problems. Specifically, the homework in the traditional course was much more often relatively complex with multiple computational steps, as the homework was assigned after the relevant lectures. In contrast, 67% of the homework problems in the IE were simpler because these problems were assigned before the relevant class. This kind of more complex, multistep problem (problem 3) showed greater gains after traditional than IE instruction. These results are in line with the theoretical and empirical work in the basic cognitive and learning science literatures, as well as research specifically on physics problem solving: transfer from training (instructed and homework) problems to test problems becomes more robust to the extent that the problems reflect more overlap [17–19,37,38]. In sum, the instructional method appeared to have less bearing on gains in quantitative problem solving than did the kinds of problems emphasized and practiced in the courses and the overlap of these problems to those on the assessment.

ACKNOWLEDGMENTS

Rebecca Trousil designed the problem-solving assessment, an initial scoring rubric, and administered the data collection. We are grateful for her efforts in this regard and for her thoughtful comments on a previous version of this manuscript. We also thank Pat Gibbons and Jose Mestre for their helpful suggestions on the manuscript. We thank Mary Stewart for assisting with blinding the set of problem-solving assessments for scoring purposes. Finally, we are

³Note that it might have been possible that the IE and traditional courses stressed different approaches to work-energy type problems: conservation of energy (a “one-step” method) versus the work-energy theorem (a “three-step” method). If so, this difference might have impacted the observed quantitative problem-solving results. However, the instruction on work-energy was similar in both courses. The IE and traditional classes emphasized the one-step method (conservation of energy), with the three-step method demonstrated in the textbooks in both courses with additional problem examples.

appreciative to the students and instructors (including R. Trousil) of the introductory physics classes at Washington University (Fall, 2009).

APPENDIX: QUANTITATIVE PROBLEM-SOLVING ASSESSMENTS

1. Problem 1

As part of the San Francisco police department's forensic team, you are investigating a potential crime scene where a pedestrian was struck by a concrete cylinder that originated from a construction site at the top of a 45 m high hill. The investigation reveals that the cylinder was moving at 40 m/s when it struck a pedestrian. Some quick measurements reveal that the thin-walled cylinder is 3.2 m long, has an outer radius of 2.0 m, and a wall thickness of 5.0 cm. Just to cover all the bases, you look up the density of concrete and note it is about 2800 kg/m³. In your report, will you conclude that this was an accident that resulted when the cylinder was jostled loose or will you conclude that this was a premeditated attempt on the person's life?

- Briefly explain with words how you would use physics to answer this question. Be sure to include key physics principles you plan to use and any assumptions or approximations that you will make.
- Now execute your plan of attack and determine conclusively whether or not this was an accident. Show your work and explain your reasoning.
- Considering the magnitude, sign, and units of your result in part (b), discuss the plausibility of your result.

2. Problem 2

In a volcanic eruption, a 2.0 kg piece of porous rock is thrown straight upward with an initial speed of 40 m/s. It travels upward a distance of 50 m before it begins to fall back to earth. If the increase in thermal energy due to air resistance on the way down is 70% of that on the way up, determine the speed of the rock when it returns to its original position.

- Briefly explain with words how you plan to solve this problem. Be sure to include key physics principles you plan to use and any assumptions or approximations that you will make.
- Show your work and explain your reasoning.
- Identical to Problem 1.

3. Problem 3

A block is released from rest at a height d and slides down a frictionless ramp and onto a first plateau, which has a length d and where the coefficient of kinetic friction is 0.50. If the block is still moving, it then slides down a second frictionless ramp through a height of $d/2$ and onto a lower plateau, which has a length of $d/2$ and where the coefficient of kinetic friction is again 0.50. If the block is still moving, it then slides up a frictionless ramp until it momentarily stops. Use energy concepts to determine where the block ultimately comes to rest. If it stops on a plateau, specify which one and the distance from the left edge of the plateau. If it stops (momentarily) on the ramp, specify its height above the lower plateau when it comes to rest.

- (a), (b), and (c) are identical to those in Problem 2.

-
- M. J. Cahill, K. M. Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey, Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020101 (2014).
 - L. Deslauriers, E. Schelew, and C. Wieman, Improved learning in a large-enrollment physics class, *Science* **332**, 862 (2011).
 - R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
 - S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. USA* **111**, 8410 (2014).
 - M. A. Ruiz-Primo, D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard, Impact of undergraduate science course innovations on learning, *Science* **331**, 1269 (2011).
 - P. Heller, T. Foster, and K. Heller, Cooperative group problem solving laboratories for introductory classes, *AIP Conf. Proc.* **399**, 913 (1997).
 - C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer Instruction: Engaging students one- on-one, all at once, *Research-Based Reform of University Physics* **1**, 40 (2007).
 - B. H. Ross, J. P. Mestre, and J. L. Docktor, Understanding how to teach physics understanding, in *Integrating Cognitive Science with Innovative Teaching in STEM Disciplines* edited by M. A. McDaniel, R. Frey, S. Fitzpatrick, and H. L. Roediger (Washington University Libraries, St. Louis, MO, 2014).
 - V. Talanquer and J. Pollard, Let's teach how we think instead of what we know, *Chem. Educ. Res. Pract.* **11**, 74 (2010).
 - K. K. Cheng, B. A. Thacker, R. L. Cardenas, and C. Crouch, Using an online homework system enhances students' learning of physics concepts in an introductory physics course, *Am. J. Phys.* **72**, 1447 (2004).

- [11] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
- [12] B. Thacker, K. Eunsook, K. Trefz, and S. M. Lea, Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses, *Am. J. Phys.* **62**, 627 (1994).
- [13] C. Hoellwarth, M. J. Moelter, and R. D. Knight, A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms, *Am. J. Phys.* **73**, 459 (2005).
- [14] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory, *Phys. Teach.* **30**, 141 (1992).
- [15] J. Larkin, J. McDermott, D. P. Simon, and H. A. Simon, Expert and novice performance in solving physics problems, *Science* **208**, 1335 (1980).
- [16] A. Van Heuvelen, Overview, case study physics, *Am. J. Phys.* **59**, 898 (1991).
- [17] M. L. Gick and K. J. Holyoak, Schema induction and analogical transfer 1, *Cogn. Psychol.* **15**, 1 (1983).
- [18] L. R. Novick, Analogical transfer, problem similarity, and expertise, *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 510 (1988).
- [19] E. L. Thorndike and R. S. Woodworth, The influence of the improvement in one mental function upon the efficiency of other functions, *Psychol. Rev.* **8**, 384 (1901).
- [20] M. A. McDaniel, A. Friedman, and L. E. Bourne Jr., Remembering the levels of information in words, *Mem. Cogn.* **6**, 156 (1978).
- [21] H. L. Roediger III., D. A. Gallo, and L. Geraci, Processing approaches to cognition: The impetus from the levels-of-processing framework, *Memory* **10**, 319 (2002).
- [22] A. K. Thomas and M. A. McDaniel, The negative cascade of incongruent generative study-test processing in memory and metacomprehension, *Mem. Cogn.* **35**, 668 (2007).
- [23] H. D. Young and R. A. Freedman, *Sears and Zemansky's University Physics with Modern Physics* 12th ed. (Pearson, New York, 2007).
- [24] T. Moore, *Six Ideas That Shaped Physics*, 2nd ed. (McGraw-Hill, New York, 2007).
- [25] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 010130 (2016).
- [26] G. M. Bodner and D. S. Domin, Mental models: The role of representations in problem solving in chemistry, *U. Chem. Ed.* **4**, 24 (2000).
- [27] V. Cahyadi, The effect of interactive engagement teaching on student understanding of introductory physics at the faculty of engineering, University of Surabaya, Indonesia, *Higher Educ. Res. Dev.* **23**, 455 (2004).
- [28] J. D. Marx and K. Cummings, Normalized change, *Am. J. Phys.* **75**, 87 (2007).
- [29] R. Ploetzner and K. VanLehn, The acquisition of qualitative physics knowledge during textbook-based physics training, *Cognit. Instr.* **15**, 169 (1997).
- [30] M. T. H. Chi, R. Glaser, and E. Rees, Expertise in problem solving, in *Advances in the Psychology of Human Intelligence*, Vol. 1, edited by R. J. Sternberg (Lawrence Erlbaum Associates, Hillsdale, NJ, 1982).
- [31] K. VanLehn and B. van de Sande, Expertise in elementary physics, and how to acquire it, in *The Development of Professional Performance: Toward Measurement of Expert Performance and Design of Optimal Learning Environments*, edited by K. A. Ericsson (Cambridge University Press, Cambridge, England, 2009).
- [32] R. Dufresne, W. J. Gerace, P. T. Hardiman, and J. P. Mestre, Constraining novices to perform expert-like problem analyses: Effects on schema acquisition, *J. Learn. Sci.* **2**, 307 (1992).
- [33] J. P. Mestre, R. Dufresne, W. J. Gerace, P. T. Hardiman, and J. S. Trouger, Promoting skilled problem solving behavior among beginning physics students, *J. Res. Sci. Teach.* **30**, 303 (1993).
- [34] D. J. Jones, K. W. Madison, and C. E. Wieman, Transforming a fourth year modern optics course using a deliberate practice framework, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020108 (2015).
- [35] T. Stelzer, G. Gladding, J. P. Mestre, and D. T. Brooks, Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content, *Am. J. Phys.* **77**, 184 (2009).
- [36] T. Stelzer, D. T. Brookes, G. Gladding, and J. P. Mestre, Impact of multimedia learning modules on an introductory course on electricity and magnetism, *Am. J. Phys.* **78**, 755 (2010).
- [37] M. A. McDaniel and M. S. Schlager, Discovery learning and transfer of problem solving skills, *Cognit. Instr.* **7**, 129 (1990).
- [38] D. T. Brookes, B. H. Ross, and J. P. Mestre, Specificity, transfer, and the development of expertise, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010105 (2011).